



非負値行列因子分解アルゴリズムに基づくメッセージ特徴の選択手法に関する研究

著者	輪島 幸治
発行年	2019
学位授与大学	筑波大学 (University of Tsukuba)
学位授与年度	2019
報告番号	12102甲第9238号
URL	http://hdl.handle.net/2241/00158138

氏 名	輪島 幸治
学 位 の 種 類	博 士 (情 報 学)
学 位 記 番 号	博 甲 第 9 2 3 8 号
学位授与年月日	平成31年4月30日
学位授与の要件	学位規則第4条第1項該当
審 査 研 究 科	図書館情報メディア研究科
学位論文題目	非負値行列因子分解アルゴリズムに基づくメッセージ特徴の選択手法に関する研究
主 査	筑波大学 教授 博士 (工学) 佐藤 哲司
副 査	筑波大学 教授 博士 (工学) 森嶋 厚行
副 査	筑波大学 教授 博士 (教育学) 芳鐘 冬樹
副 査	筑波大学 准教授 博士 (情報学) 手塚 太郎
副 査	筑波大学 教授 博士 (工学) 天笠 俊之

論 文 の 要 旨 (2,000 字程度)

近年、メールやチャット、ソーシャルネットワークサービス(SNS)など、テキストによる情報交換・情報共有の機会が増えている。また、商業化されたマス・コミュニケーションと個人間のパーソナル・コミュニケーションの垣根が低くなり、特定個人に向けたメッセージが公開されることで、プライバシー侵害や炎上を誘起する事例も増えてきている。

本論文は、このような社会的背景を踏まえて、不特定多数のメッセージ受信者が好意的に受け止める、あるいは、プライバシー侵害にあたらないと判断されるテキストの特徴を明らかにすることを目的として、特徴次元の圧縮に優れた非負値行列因子分解アルゴリズムを用いてメッセージ特徴を選択する手法を6章構成で論じている。

第1章は、本研究の背景を述べた章である。最初に、携帯型端末の小型化と高性能化とともに登場したスマートフォンは、コミュニケーション手段を電話からメールやチャットなどのテキストへと移行させてきたこと、この変化に伴って、これまでマスメディアが主役を担っていた情報発信は、個人やコミュニティへと開放され一個人でも不特定多数に情報を発信できる環境が整ってきたなど、情報化社会の変容について述べ、このような個人による情報発信に基づいて形成されるメディアはソーシャルメディアと呼ばれ、21世紀を特徴付けるコミュニケーションメディアとなっていると結論づけている。

次に、ソーシャルメディアでは、従来のマスメディアが編集というプロセスを経て、選別された情報を発信してきたモデルとは大きく異なり、情報の発信者から直接かつ大量に発信される傾向が強くなっていること、この結果、情報の受信者は、不要と判断する情報が多い伝播メディアを遮断し、以降は当該メディアからの情報を一切閲覧しなくなるという、受信者の極端な行動が円滑なコミュニケーションを阻害する要因であると指摘している。また、情報に対する受信者の

反応も多様化しており、受信者の中には過剰な反応や炎上のきっかけを作る者も少なからず存在することを示し、このような事象の発生を情報の発信者が予見することは困難であり、個々の受信者の都合に合わせた情報発信は一層難しくなっていることも指摘している。

以上述べた社会的な背景とそこから生じる課題を踏まえて、本論文では、記事の閲覧数から判断される話題性および共感を呼ぶメッセージ特徴、公的な出版物である「環境白書」とその啓発書である「こども環境白書」を比較した平易化テキストのメッセージ特徴、および、記事の閲覧者がプライバシー侵害と判断する記事のメッセージ特徴の選択手法を提案するとしている。

第 2 章では、本論文で着目する 4 種類のメッセージ特徴、すなわち、話題性、共感、平易化、プライバシーに関する関連研究を概観し、本論文の位置づけを明らかにしている。話題性および共感に関する関連研究では、トレンドキーワードや情報カスケードなどの既存手法を概説し、研究トレンドを明らかにしている。テキスト情報の平易化に関する関連研究では、コンテンツの難易を判別する研究とコンテンツを平易化する研究を概観している。プライバシーに関する関連研究では、プライバシーの多義性やプライバシー侵害などに関する既存研究を紹介し、これまでの動向を明らかにしている。

第 3 章は、メッセージ特徴とその選択手法を提案する章であり、メッセージ特徴を選択するための非負値行列因子分解アルゴリズムを示し、本論文で定義するメッセージ特徴を論じている。一般に、機械学習では、特徴量が多いと教師データに過学習をすることが知られており、特徴量の選別が汎化能力を向上することが知られている。本論文では、受信者に好意的に受け入れられるメッセージの特徴を、非負値行列因子分解アルゴリズムの基底として分解し、メッセージ特徴として有効な基底を、分類精度や推定精度によって選択する。このために使用する分類器、および、非線形回帰、グラフィカルモデルについても詳述している。具体的に用いた分類器は、Ada Boost、Random Forests、MLP (Multi-Layer Perceptron)、K-Nearest Neighbors (K-NN) である。分類精度の評価指標は、一般的に広く用いられている適合率、再現率、F-値である。また、推定に使用する非線形回帰手法としてサポートベクタ回帰モデルも述べている。推定の精度は、目的変数に対する予測誤差である MAE (Mean Absolute Error) と RMSE (Root Mean Squared Error) の 2 つで評価している。グラフィカルモデルは、本研究において因果関係の推定に使用しており、広く使われているベイジアンネットワークについて述べている。

第 4 章および第 5 章は、提案手法の有効性を示すために行った実験・評価の章である。第 4 章では、グレゴリー・ベイトソンの情報の定義である「“違い”を生む“違い”」を根拠として、差異に基づくメッセージ特徴の選択を行っている。差異とは、正解と不正解の明確な分離ができない学習データに対して、比較する 2 つの文書集合における係数値の差異であり、差異に着目することで、メッセージ特徴を評価できるとしている。質問記事における閲覧数の差異に着目した話題性の判別実験と、普及啓発書「こども環境白書」と平易化前の「環境白書」とを比較する平易化の判別実験を行い、提案手法の有効性が確認できたとしている。

第 5 章では、回帰に基づいてメッセージ特徴の選択実験を行っている。サポートベクタ回帰モデルを用いて非線形回帰を行い、プライバシー侵害および共感を目的変数としてメッセージ特徴を評価している。プライバシー侵害の懸念の有無をラベル付けした SNS 投稿記事と、返信数が異なる質問記事を対象とした分類実験とから、提案手法の有効性が確認できたとしている。

第 6 章では、本研究における結論を述べている。提案手法および評価実験の結果を総括し、残された課題、および、成果を踏まえた望ましい発展の方向について、私見を交えた展望も示している。

審 査 の 要 旨 (2,000 字以上)

【批評】

情報化社会の進展に伴って、コミュニケーション手段の主役は電話からメールやチャットなどのテキストへと移行してきた。特に、常時携帯できるスマートフォンを利用することで、個人が不特定多数の受信者に大量かつ多様な情報を容易に発信できる。このような個人間の情報発信、情報共有を主体とするソーシャルメディアは、様々な伝播メディアの集合体として形成されており、情報の発信者および受信者は、自身に都合が良く快適な伝播メディアを選択的に利用し、不要あるいは不快な情報が多い伝播メディアは遮断してしまう。また、受信者の中には過剰な反応や炎上のきっかけを作るものも少なくないなど、ソーシャルメディアが持続可能で安定したコミュニケーション手段となるには多くの課題があることは広く知られている。

本研究は、このような社会的課題の解決を目的に、発信された情報すなわちテキストメッセージを入力として、多様な受信者に幅広く受け入れられるか、過剰な反応や炎上を引き起こす懸念が無いかなど、メッセージの高次な特徴を選択・評価する手法を提案している。ここで設定された課題は、ソーシャルメディアにおける本質的な課題であり学術的な価値も高いと言える。また、4 種類の具体的なコミュニケーションの場を設定して、著者が好ましいと設定したコミュニケーションとするためのメッセージ特徴を、実践的に明らかにしようとする取り組みは、社会的にも意義のある研究であると言える。

第1章では、ソーシャルメディアが成長してきた社会的背景を述べた後に、不特定多数の者を対象とするマス・コミュニケーションの 3 要素: 事実の報道、解説・啓発、娯楽に加えて、近年社会問題となっているプライバシー侵害を合わせた 4 要素に基づいて、具体的に取り組むコミュニケーション場の導出を試みている。論文にはやや冗長な説明が含まれているが、実験評価を行っている 4 種類のコミュニケーション場を取り上げる動機は十分に理解できる。

第 2 章は、本研究で取り組む 4 種類のコミュニケーション場に関する先行研究と、本研究が目指すメッセージ特徴の選択に関する既知のテキスト特徴量および特徴量選択手法について述べている。本研究の視座から関連研究を概観するという観点からの整理に課題は残るが、本研究に関連する 100 以上の先行研究を調査していることから、関連研究の章としては十分な内容となっていると言える。

第 3 章は、メッセージ特徴の選択手法と選択された特徴の有効性を判別する手法を提案する章であり、本論文の中核をなす章となっている。メッセージ特徴を選択するための非負値行列因子分解アルゴリズム(NMF)、および、本研究で定義するメッセージ特徴を論じた後に、有効性判別に使用する分類器、および、非線形回帰、グラフィカルモデル等について詳述している。数式や図表を用いて分かりやすく説明しようとしていることは評価できる。非負値行列因子分解アルゴリズムの採用に際して、既存研究で優れていることが示されていることだけでなく、本研究の目的に照らして妥当あるいは優位である根拠を示すと、より説得力のある論文になると期待される。

第 4 章は、差異に基づくメッセージ特徴の選択手法を提案している。グレゴリー・ベイトソンの情報の定義である「“違い”を生む“違い”」に触発されて、正解と不正解の明確な学習データが得られない場合でも係数値の差異に着目することで、メッセージ特徴を評価できるとしている。記事の話題性は閲覧数と正の相関があることは直感的に理解できるが、閾値を設けて二値分

類することは難しい。また、行政が出版している白書と子供向け啓発書の比較においても、啓発書の方が平易化された文書であることは間違いないが、どこがどの程度に平易化されているかを切り分けることは困難である。このような課題に対して、差異に基づくメッセージ特徴の選択手法は有効に機能したと述べている。しかしながら、子供向け啓発書は、表現の平易化だけでなく、取り扱う話題も取捨選択されていることから、この実験で論じることができる有効性の範囲は限定的であり、今後、他の文書集合を用いた実験を行うなどすることで、より精緻な議論が行えるであろう。本論文は、そのための橋頭堡としての価値が認められる。

第 5 章は、回帰に基づくメッセージ特徴の選択手法を提案している。サポートベクタ回帰モデルを用いて非線形回帰を行い、プライバシー侵害および共感を目的変数としてメッセージ特徴を評価している。論文でも述べているように、どのような文章がプライバシー侵害に該当するかは難しい問題である。実験では、記事を閲覧した実験参加者が侵害していると感じるかという主観的な評価に基づいて分類実験を行い、実験参加者と同程度の分類ができたことをもって、提案手法の有効性を論じている。侵害が疑われる記事 96 件を 55 名の実験参加者がラベル付けした正解データで提案手法の有効性は評価できているが、記事によって実験参加者間の一致度は異なると思われることから、カッパ係数などの一致度評価と組み合わせ、人間による評価と提案手法の評価に見られる差異を詳細に分析することで、更に有益な知見が得られる、研究の発展性を感じる。

第 6 章は、本研究の総括と今後の展望を述べている。最後に記述されている展望はやや唐突な内容であることは否めないが、本研究の成果・到達点を橋頭堡として外挿できる研究の発展は幅広く、得られた知見に新規性があり、学術的な価値も高い論文となっている。

以上のように、本研究は社会的な背景を踏まえた妥当な問題設定であり、提案手法の新規性、有効性は十分に評価できる内容といえる。評価実験には、荒削りな面も散見されるが、研究の学術的価値を否定するものではなく、むしろ、新しい課題に挑戦していることの証左であるとも言える。

これらを総合的に判断すると、本論文は情報学の学位論文として十分な内容を有すると認められる。

【最終試験結果】

平成 31 年 3 月 27 日、図書館情報メディア学位論文審査委員会において、審査委員全員出席のもと、本論文について著者に説明を求めた後、関連事項について質疑応答を行った。引き続き、「図書館情報メディア研究科博士後期課程（課程博士）の学位論文審査に関する内規」第 23 項第 3 号に基づく最終試験を行い、審議の結果、審査委員全員一致で合格と判断された。

【結論】

よって、著者は博士（情報学）の学位を受けるに十分な資格を有するものと認める。